HHS Public Access

Author manuscript

Cell. Author manuscript; available in PMC 2017 July 14.

Published in final edited form as:

Cell. 2016 July 14; 166(2): 343–357. doi:10.1016/j.cell.2016.05.072.

An abundant class of non-coding DNA can prevent stochastic gene silencing in the *C. elegans* germline

Christian Frøkjær-Jensen[1,2,3], **Nimit Jain**[4], **Loren Hansen**[2], **M Wayne Davis**[1], **Yongbin Li**[8], **Di Zhao**[8], **Karine Rebora**[6], **Jonathan RM Millet**[6], **Xiao Liu**[5,8], **Stuart K Kim**[5,7], **Denis Dupuy**[6], **Erik M Jorgensen**[1], and **Andrew Z Fire**[2,7]

[1]Howard Hughes Medical Institute, Department of Biology, University of Utah, Salt Lake City, UT 84112, USA

[2]Department of Pathology, Stanford University, Stanford, CA 94305, USA

[3]Department of Biomedical Sciences and Danish National Research Foundation Centre for Cardiac Arrhythmia, University of Copenhagen, Copenhagen, Denmark

[4]Department of Bioengineering, Stanford University School of Medicine, Stanford, CA 94305, USA

[5]Department of Developmental Biology, Stanford University Medical Center, Stanford, CA 94305, USA

[6]IECB, University of Bordeaux, laboratoire ARNA-INSERM, U869, F-33600 Pessac, France

[7]Department of Genetics, Stanford University, Stanford, CA 94305, USA

[8]School of Life Sciences, Tsinghua University, Beijing 100084, China

Summary

Cells benefit from silencing foreign genetic elements but must simultaneously avoid inactivating endogenous genes. Although chromatin modifications and RNAs contribute to maintenance of silenced states, the establishment of silenced regions will inevitably reflect underlying DNA sequence and/or structure. Here we demonstrate that a pervasive non-coding DNA feature in *Caenorhabditis elegans*, characterized by 10-basepair periodic A_n/T_n-clusters (PATCs), can license transgenes for germline expression within repressive chromatin domains. Transgenes containing natural or synthetic PATCs are resistant to position effect variegation and stochastic silencing in the germline. Among endogenous genes, intron length and PATC-character undergo dramatic changes as orthologs move from active to repressive chromatin over evolutionary time, indicating

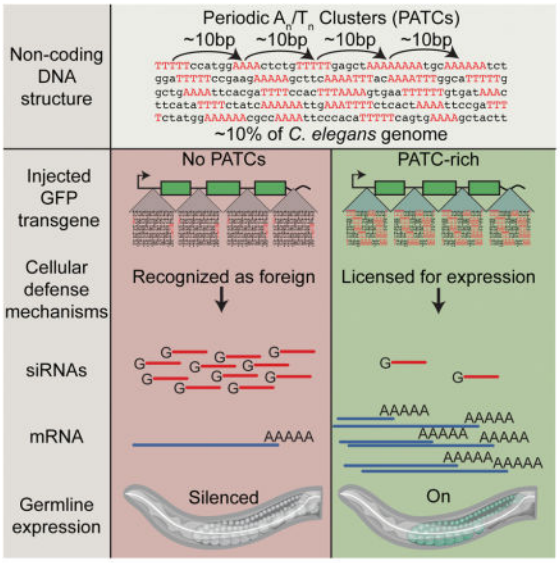
Correspondence: jorgensen@biology.utah.edu (E.M.J), afire@stanford.edu (A.Z.F).

Author contributions

Conceptualization, C.F.J. and A.Z.F.; Software, L.H., M.W.D. and A.Z.F.; Formal Analysis, C.F.J. and A.Z.F.; Investigation, C.F.J., N.J., L.H., Y.L., D.Z., K.R., J.R.M.M., X.L., D.D.; Writing - Original Draft, C.F.J.; Writing - Review & Editing, C.F.J., E.M.J., A.Z.F.; Supervision, C.F.J., X.L., S.K.K., D.D., E.M.J., A.Z.F.; Funding Acquisition, C.F.J., E.M.J., A.Z.F.

a dynamic character to the A_n/T_n periodicity. We propose that PATCs form the basis of a cellular immune system, identifying certain endogenous genes in heterochromatic contexts as privileged while foreign DNA can be suppressed with no requirement for a cellular memory of prior exposure.

Graphical Abstract



eTOC Paragraph

A non-coding DNA sequence pattern can prevent epigenetic silencing of transgenes in germ cells, revealing a functional role for an abundant class of non-coding DNA and a possible mechanism by which cells may recognize and silence foreign DNA

Introduction

Invasive DNA derived from viruses, retrotransposons, and DNA transposons constitute a substantial challenge to organisms. Uncontrolled replication of transposable elements will compromise the host's genome (Malone and Hannon, 2009) and consequently, cellular defense mechanisms have evolved to detect and silence foreign DNA. These mechanisms are particularly well developed in germ cells, where deleterious changes will impact the fitness of subsequent generations. In eukaryotes, several classes of small RNAs (~20–30 nucleotides long) form complexes with Argonaute proteins to silence foreign nucleic acids by degrading target mRNAs (Zamore et al., 2000) and by transcriptional silencing via RNA-directed heterochromatin formation (Volpe et al., 2002). In germ cells, genome surveillance is mediated in part by a large class of small RNAs (piRNAs) that interact with Argonautes from the Piwi clade (e.g. Aravin et al., 2007; Batista et al., 2008; Das et al., 2008).

Small RNAs act as a recognition system for transcription from invasive DNA, but introduce the danger of silencing endogenous genes. In *C. elegans*, several potentially-related mechanisms have been proposed to protect endogenous genes against silencing. At a

chromatin level, activating H3K36 histone marks deposited on germline-expressed genes provide a positive feed-forward mechanism promoting expression in subsequent generations (Rechtsteiner et al., 2010). Operating alongside and/or in parallel, RNA-based protection systems have been shown to use maternal RNAs to license forthcoming expression of genes (Johnson and Spence, 2011). A licensing mechanism based on small RNAs associated with the CSR-1 Argonaute protein has been proposed to protect endogenous genes from piRNA-mediated silencing (Shirayama et al., 2012). Notably, these mechanisms would only propagate germline expression decisions made in prior generations. What other features, such as genome position or DNA sequences, determine initial licensing for germline expression has remained an open question. Specifically, how can a gene be selected (or rejected) for expression if there is no chromatin mark or small RNA population to indicate whether the gene was expressed in prior generations? To probe this question, we focused on *de novo* expression and silencing of transgenes in the germline of *C. elegans*.

Transgenes have been a useful tool to determine the effects of large-scale genome organization in several organisms, most notably *Drosophila* (Elgin and Reuter, 2013). In *C. elegans*, transgenes are notoriously difficult to express in the germline, with rapid silencing of episomal DNA (Kelly et al., 1997) and progressive silencing of many single-copy genomic transgenes (e.g., Shirayama et al., 2012). Here we study a stochastic process in which some single-copy insertions variegate in somatic cells and are frequently silenced in germ cells. Variegation and silencing mirror a chromosomal pattern that corresponds to the organization of the genome into broad domains (e.g. Liu et al., 2010). We find that a non-coding DNA structure, called Periodic A_n/T_n Clusters (PATCs) (Fire et al., 2006) can license transgenes for expression in the germline. Thus, PATCs constitute an abundant class (comprising ~10% of the *C. elegans* genome) of functionally important non-coding DNA in nematodes that may safeguard endogenous genes from silencing in repressive chromatin environments. We propose that lack of PATCs in foreign DNA may be one characteristic used by nematodes to silence invasive genetic material.

Results

Germline expression is sensitive to large-scale genome organization

To explore transcriptionally permissive and repressive genomic regions for germline expression we inserted a ubiquitously expressed transgene (P*dpy-30*:GFP:H2B) into random locations by transposition (Frøkjær-Jensen et al., 2014) (Figure 1A, Table S1). Of 67 insertion strains analyzed, all had visible somatic expression but most (51 strains) had limited germline expression. We categorized germline expression from animals maintained at 25°C into three classes: complete germline expression (mitosis, early, and late meiosis), early germline silencing (fluorescence only visible in late meiosis), and full germline silencing (no visible expression in any germ cells) (Figure 1A, Figure S1). Additionally, some strains showed transgene variegation and were categorized as “variable”.

Strains with early or full germline silencing appeared to cluster in non-random chromosomal patterns (Figure 1B). First, all insertions into the X chromosome showed early germline silencing (Figure S1). This pattern is consistent with broad inactivation of the X chromosome in the early meiotic germline by homologs of the Polycomb Repressive

Complex 2 (PRC2) (Fong et al., 2002) and de-repression at late meiotic stages (Kelly et al., 2002). Second, 15 0f 25 strains with transgenes inserted within the central 50% of autosomes were expressed, whereas only 1 of 30 transgenes inserted into the distal 25% of autosome arms were expressed in the germline ($P < 0.001$, Fisher's exact test) (Figure 1B). This pattern of expression is in agreement with the observed central and distal autosomal domains based on recombination frequency (Brenner, 1974; Rockman and Kruglyak, 2009), histone modifications (Gu and Fire, 2010; Liu et al., 2010), and heterochromatin protein 1 (HP1) distribution (Garrigues et al., 2015). Targeted insertion of a P*dpy-30*::GFP transgene (Frøkjær-Jensen et al., 2008, 2014) showed a similar pattern of germline expression and frequent position-dependent stochastic silencing on autosome arms (Figure S1).

Transgene expression in the soma is position-dependent and variegates

Although we were primarily interested in germline-specific silencing, we also tested whether somatic transgene expression was position-dependent. We generated random insertions with a strong ubiquitous promoter *eft-3* (*eef-1A.1*) that expressed a bright red fluorophore (*tdTomato*) and screened transgenic animals for low expression at the L1 stage. Qualitatively, most strains showed reproducibly bright expression and we were unable to isolate any animals with fully silenced transgenes. However, a subset of strains (40 of ~800 insertions) showed "mottled" expression in a limited number of somatic cells (Figure 2A). We scored the fluorescence qualitatively at the L1 stage by assigning animals to five expression classes (Class 1 (brightest) to Class 5 (most silenced)) (Figure 2A, Table S2).

Insertions into chromosome centers generally belonged to the two brightest classes whereas most dim transgenes were inserted into distal regions (Figure 2D). Genes silenced in the L1 stage corresponded to local genomic environments (± 1 kb) previously found to be significantly enriched in repressive H3K9me3 histone marks at a later larval (L3) stage (Liu et al., 2010) (Figure 2E). We generated an H3K9 methylation dataset from L1 animals and verified this association between transgene expression and H3K9me3 histone marks (Figure 2F). Finally, silenced insertions were frequently inserted into chromatin identified as repressive based on global chromatin marks at the L3 stage (Polycomb repressed and heterochromatin states, hiHMM 10–13) (Ho et al., 2014) (Figure 2G, Figure S2C).

We also measured transgene variegation with quantitative imaging techniques on strains from Class 1, Class 4, and Class 5. One method automatically identified and assigned fluorescence levels to 363 of the 558 cells from fixed L1 animals (Liu et al., 2009). L1 imaging detected expression in significantly more cells and higher levels of expression in the Class 1 strain compared to Class 4 and Class 5 strains (Figure 2B, Figure S2A) and a higher degree of variation in dim strains compared to bright strains was evident (e.g. E lineage, Figure S2B). A second technique relied on imaging live worms across different developmental stages by flow cytometry (Dupuy et al., 2007), which confirmed higher transgene expression in Class 1 compared to Class 4 and Class 5 at the L1 stage (EG7213 = 19.2 ± 0.4, EG7207 = 11.3 ± 0.2, EG7209 = 7.9 ± 0.1, mean ± SEM) (Figure 2C). At later larval stages, the two Class 1 and Class 4 strains increased in absolute fluorescence, whereas the one Class 5 strain remained mostly dim throughout development (Figure 2C). Thus,

transgene silencing in early larval stages is not necessarily a permanent state, possibly tracking developmental changes in local chromatin environments (Meister et al., 2010).

In sum, we observed evidence for transgene variegation and an influence of large-scale genome domains on somatic expression. At the same time, only a small subset of insertions variegated. These data extend the longstanding observation that it is easier to express transgenes in the soma compared to the germline of *C. elegans* (e.g., Kelly et al., 1997).

Transgene silencing in the germline depends on promoter elements

The silencing character of any given chromosomal region might conceivably be universal (so that any insertion would be shut down) or insertion-dependent (capable of silencing some transgenes but not others). We examined whether position-dependent germline silencing was transgene-dependent by using two germline-specific promoters: P*mex-5* and P*pie-1*. Similar to P*dpy-30*:GFP (Figure 1B), P*mex-5*:GFP and P*pie-1*:GFP transgenes were mostly silenced on the X-chromosome and frequently silenced on autosome arms compared to centers (GFP positive: P*mex-5*, 11 of 17 arm, 25 of 25 center, $P < 0.01$; P*pie-1*: 28 of 48 arm, 34 of 37 center, $P < 0.01$; Fischer's test) (Figure 3A). Also, P*pie-1*:GFP insertions were more frequently silenced in genomic regions with high levels of H3K9me3 (early embryo, EE) (Liu et al., 2010) (Figure 3B) and in regions that immunoprecipitate with a nuclear lamina protein (LEM-2) (Ikegami et al., 2010) (Figure 3C). However, the *mex-5* and *pie-1* promoters were significantly more active compared to P*dpy-30* from autosome arms (GFP-positive insertions: P*mex-5*, 11 of 17 arm; P*pie-1*: 28 of 48 arm; P*dpy-30*: 2 of 31 arm, $P < 0.01$, Fischer's, Figure 3C) and from repressive hiHMM chromatin states (Ho et al., 2014) (Figure S3F).

Targeted insertions of P*pie-1*:GFP showed a similar pattern of germline expression and stochastic silencing (GFP-positive insertions: 13 of 18 center, 0 of 24 arm, $P < 0.01$, Fischer's) (Figure 3E). One insertion site (*oxTi176*) near the transition between domains of low and high H3K9me3 showed highly variable P*pie-1*:GFP (Figure 3E) and P*mex-5*:GFP (Figures S3I) expression; it is possible that insertions into chromosomal locations bordering heterochromatin are particularly prone to stochastic silencing and position effect variegation, similar to what has been observed in *Drosophila* (Elgin and Reuter, 2013).

These data provide support for genome position as a strong determinant of germline expression and suggest that some germline promoters are more resistant to silencing imposed by heterochromatic domains.

A transgene rich in periodic A/T clusters is expressed from repressive chromatin domains

To resolve how endogenous genes are protected from stochastic silencing in the germline, we looked for a common DNA character that might safeguard genes from the surrounding repressive heterochromatin environment. Periodic A_n/T_n clusters (PATCs) are an abundant class of non-coding DNA that are enriched in germline expressed genes on autosome arms (Fire et al., 2006) and are anti-correlated with H3K9 methylation (Gu and Fire, 2010). PATCs are composed of short clusters of adenines and thymines spaced approximately 10 basepairs apart, thereby "coating" one face of the DNA helix with A_n/T_n clusters over

extended runs (Figure 3F). It is notable that the relatively silencing-resistant *pie-1* and *mex-5* promoters contain PATCs (Figure 3A) whereas P*dpy-30* does not (not shown). We confirmed the previously established positive association between germline expression and PATCs (Fire et al., 2006) with more recent gene models (WS245) and gene expression profiles from isolated germlines and single-cell oocytes (Ortiz et al., 2014; Stoeckius et al., 2014; Wang et al., 2009) (Figure 3G, Figure S3).

As a starting point for investigating the role of PATCs, we generated random insertions of a *gfp*-tagged *smu-1* gene (*smu-1:gfp*). *smu-1* has moderately strong overall PATC content (Figure 3G) with PATCs distributed across the endogenous promoter, gene body, and 3′ UTR (Figure 3A). *smu-1*, and the related gene *smu-2*, have the highly unusual property that simple, extrachromosomal arrays (hereditary, highly repetitive episomal DNA structures) with these genes are readily expressed in both soma and germ cells (Spartz et al., 2004; Spike et al., 2001). When inserted randomly, all *smu-1:gfp* insertions were expressed in somatic cells and in the germline when cultured at 25°C (with one exception, which is likely a damaged transgene insertion). Notably, despite some initial silencing, X-linked *smu-1*:GFP transgene inserts were also expressed in the full germline after propagation for 3–4 generations (Figure 3A). Germline de-silencing over time is not a general feature of transgenes; X-linked P*mex-5* insertions remained silenced over the same number of generations (5 of 5 P*mex-5* strains, Table S2). These observations are unlikely to be be explained by relative promoter strength: *smu-1* expression is generally low compared to *mex-5*, *pie-1*, and *dpy-30* expression, as measured by endogenous gene expression (Stoeckius et al., 2014) or visual inspection of transgene fluorescence.

If PATCs contribute to prevent silencing, then insertion into a PATC-rich chromatin environment might also promote germline expression? We analyzed the local PATC content near P*pie-1:gfp* insertions and found no positive association between high PATC-content and germline expression (Figure 3D). We also observed no association between somatic transgene expression (P*eft-3:tdTomato*) and local PATC environment (Figure S3J), suggesting that insertion into A_n/T_n clusters does not in itself protect from silencing.

These data show that at least one PATC-rich transgene (*smu-1*) is remarkably resistant to germline silencing and suggest that, if PATCs permit germline expression, only do so when they are part of the transgene itself.

PATCs in introns of *gfp* reduces stochastic germline silencing

Are PATCs sufficient to safeguard transgenes from gene silencing? To test the effect of PATCs in a consistent context we used a ubiquitous promoter (P*eft-3*) and 3′ UTR (*tbb-2*) with few PATCs. We expressed a *gfp* with minimal piRNA homology that had been optimized for high expression (Figure 4A). Only the PATC content within introns was varied; in particular, the initial 68 basepairs of *gfp* were kept invariant to minimize possible effects on translation efficiency and all intronic splice junctions were identical between transgenes to minimize possible differences in silencing caused by spliceosome stalling (Dumesic et al., 2013). A standard *gfp* and optimized *gfp*s with short synthetic introns or introns from a neuronal gene *snt-1* were frequently silenced in the germline from the center of Chr. V (*oxTi365*, insertion at 25°C) (Figure 4A). In contrast, PATC-rich introns from

smu-1, *smu-2*, or the *C. briggsae* ortholog of *smu-1* (*cbr-smu-1*) significantly reduced stochastic transgene silencing. In the repressive chromatin environment on the arm of Chr. V (*oxTi173*), we observed a similar pattern of stochastic silencing, except that all transgenes were expressed at lower frequency (Figure 4A). Similarly, silencing of a *gfp:cdk-1* transgene was reduced when fused to an optimized GFP containing PATCs (Figure S4C).

Are the intronic PATCs responsible for reduced germline silencing, or do the introns harbor other signals that increase expression, for example, germline specific enhancers? First, no single *smu-2* intron increased the frequency of germline expression (Figure S4A). Second, a *gfp* with *smu-1* introns inserted with a minimal promoter (*pes-10*) did not result in visible germline or somatic expression (data not shown). These data argue against the presence of strong enhancers in the introns. Third, we generated synthetic introns with PATCs by gene synthesis. P*eft-3:gfp* transgenes with synthetic PATCs showed partial but significant resistance to germline silencing in both permissive and repressive chromatin domains (Figure 4A). Fourth, increasing the number of synthetic PATC introns reduced stochastic gene silencing in repressive environments, suggesting an additive effect of PATCs (Figure S4B), although one of our synthetic introns consistently decreased expression. Fifth, the ability of PATC-rich *smu-1* introns or synthetic introns to prevent germline silencing was lost when we shuffled the intron sequences to eliminate the A-T clusters but maintained overall nucleotide composition (Figure 4B). This indicates that the basepair composition or specific length of these introns does not in itself improve germline expression.

By contrast, we were unable to confidently demonstrate that PATCs reduce somatic transgene variegation. We inserted P*eft-3*:tdTomato transgenes with no introns (cDNA), a codon-optimized tdTomato with short synthetic introns or introns from *smu-1* with PATCs into central (*oxTi365*) and distal (*oxTi173*) locations. Somatic expression quantified by visual classification, with flow cytometry, and with automated identification in L1 animals showed at most a very modest increase in expression from transgenes with PATCs and only at later larval stages (Figure S4D–E).

In sum, native and synthetic PATC-rich introns placed in a foreign coding region can reduce position dependent silencing in the germline.

PATC-rich introns across a range of lengths and from many genes reduce germline silencing

Introns from *smu-1* and *smu-2* could efficiently reduce transgene silencing but were derived from genes that are unusually resistant to germline silencing (Spartz et al., 2004; Spike et al., 2001). Furthermore, we maintained animals at 25°C, a temperature that empirically promotes germline expression of transgenes (Strome et al., 2001) but also reduces fecundity and is above the thermal tolerance of some *C. elegans* isolates (e.g. the Bergerac isolate) (Hirsh et al., 1976). To more fully characterize PATC introns, we investigated the role of intron length, intron diversity, and temperature on germline expression with five pairs of GFP with or without PATCs in introns. To select introns, we analyzed the PATC density of all protein-coding introns individually (112,275 introns, WS245): introns have a median length of 69 bps and a prominent peak of PATC-rich introns near ~900 bp (Figure 5A). We selected 30 introns spanning lengths from ~150 bp to ~900 bp that were derived from 29

different endogenous genes and inserted three introns into each P*eft-3*:gfp transgene. Each individual PATC-containing intron (solid black circles) was matched to an intron with no PATCs (open circles); the introns were matched to within 10% based on intron length and germline expression of the endogenous genes containing the two introns (Table S3).

At the central insertion site (*oxTi365)* we observed variable germline expression at 20°C from transgenes with no PATCs (Figure 5B). In contrast, all five transgenes with PATCs were expressed at high frequencies (Figure 5B). The effects of PATCs on somatic transgene expression were mixed; transgenes with PATCs were generally well expressed but only one PATC-rich transgene had significantly higher expression than a poorly expressed matched control (Figure 5B). In repressive chromatin (*oxTi173*) we observed a strong and consistent effect of PATCs: transgenes containing PATCs were less frequently silenced in the germline (4 of 5 matched transgenes) and there was no pervasive enhancement of somatic expression (Figure 5C). These differences did not generally appear to be caused by differences in piRNA homology, with one possible exception (900 bp, non-PATC) (Figure 5D). We note that only transgenes with the longest introns (700–900 bps) were expressed at high frequency in repressive chromatin and we observed a good association ($R^2 = 0.89$) between PATC content and resistance to germline silencing (Figure 5E). It is possible that longer introns or higher PATC densities are required to efficiently prevent germline silencing within highly repressive chromatin.

In sum, these data demonstrate a consistent ability for PATC-rich introns to reduce germline silencing and small effects, if any, on somatic transgene expression.

mRNA is depleted and small antisense RNAs are enriched in strains with silenced transgenes

To determine at what stage transgenes were silenced we performed single molecule RNA fluorescence *in situ* hybridization (smFISH) (Raj et al., 2008) and RNA sequencing experiments on active and silenced P*eft-3:gfp* insertions. We observed many individual *gfp* transcripts (diffraction limited cytoplasmic spots; Figure 6A) and frequent transcriptional foci (brighter nuclear spots; Figure S5) by smFISH against *gfp* in germlines with GFP expression (PD1538). In contrast, we detected few transcripts and no transcriptional foci in GFP negative germlines from wild-type animals (negative control) or from fully silenced strains (e.g. PD1540). The same was true for a transgene with PATCs that was fully silenced (PD1539) or in GFP-negative animals from a transgene that was infrequentely silenced (PD1537) (Figure 6A and Figure S5).

To examine silencing on a bulk level, we sequenced total RNA isolated from synchronized young adult hermaphrodites and observed a depletion of *gfp* mRNA in animals with silenced P*eft-3:gfp* transgenes (Figure 6B). A strain with frequent GFP expression in the germline (PD1537) had approximately 10-fold more transcripts than animals with a fully silenced PATC-rich *gfp* (PD1539) or a fully silenced *gfp* with no PATCs (PD1540). We did not capture unspliced *gfp* pre-mRNAs sequences in RNA samples from strains with active or silenced transgenes. In combination with the lack of detectable transcripts in the nucleus by smFISH, we found no evidence for accumulation of unspliced transcripts in the germ cells of animals with silenced GFPs.

To further understand the mechanisms involved in transgene silencing, we isolated and sequenced populations of small RNAs from synchronized young adult hermaphrodites (Figure 6C). Animals carrying silenced GFPs with PATCs (PD1539) or lacking PATCs (PD1540) were 10–20 fold enriched for detectable small antisense RNAs against GFP (primarily 21–23G RNAs) compared to a strain with frequent GFP expression (PD1537). Fully silenced *gfp* transgenes (i.e. full stochastic silencing from the time of insertion) with or without PATCs were indistinguishable based on the level of *gfp* mRNA and small anti-sense RNAs (PD1539 versus PD1540) (Figure 6B, C).

Thus, PATCs within introns of a foreign gene confer significant but incomplete protection from stochastic silencing, in agreement with visible germline fluorescence (Figures 4,5,6). It is possible these observations reflect a lack of PATCs in *Peft-3* and the *tbb-2* 3′ UTR; perhaps, full protection from stochastic silencing also requires PATCs in regions flanking the coding sequence, as observed for the *smu-1* transgene (Figure 3A).

In sum, these data suggest that transgenes are most likely transcriptionally silenced with the abundance of small RNAs potentially indicative of an RNAi-like mechanism maintaining, or potentially initiating, the silenced state via secondary 22G siRNAs.

Evolutionary adjustment of PATC content for different genomic environments

Genomes are under selective pressure and we expect functionally important sequence characteristics to be evolutionarily conserved in closely related species. PATCs are well conserved in the *Caenorhabditis* genus (5–11% of the total genome sequences are PATC-rich), whereas conservation is mixed in more distantly related nematodes (Figure 7A and Table S4). Outside of nematodes, most genomes do not contain comparable frequencies of PATCs, although some distantly related organisms have PATC-like structures in their genomes (e.g. the centipede *S. maritima*) (Figure 7B and Table S4).

If PATCs are functionally important, we might also predict the PATC content of genes to change in response to changes in chromatin environment, for example those caused by large-scale genome rearrangements. To test this prediction, we compared *C. elegans* and *C. briggsae*, whose most recent common ancestor existed approximately 20 million years ago (Ross et al., 2011). The overall genomic PATC-content is similar in the two species (Figure 7A) and PATCs from *cbr-smu-1* were able to safeguard transgenes in *C. elegans* (Figure 4A) suggesting functional conservation. Based on recombination frequencies (Ross et al., 2011) a chromosome structure with distinct center and arm domains is conserved in *C. briggsae,* and PATCs are similarly enriched on autosomal arms (Figure S6A). We analyzed unique *C. elegans* and *C. briggsae* ortholog pairs and determined their PATC content as a function of genomic location and expression in the germline (Figure 7C). Orthologs pairs that remain on arms have longer introns (Figure 7D) and higher PATC frequency (Figure 7E) compared to ortholog pairs that remained at a central location over evolution. Ortholog pairs that change chromatin domain show reciprocal changes: the ortholog residing in repressive chromatin has ~3 fold longer introns (Figure 7D) and ~4-fold higher PATC content (Figure 7E) than the ortholog residing at a central domain. Examples of these large changes in intron size and PATCs as a function of genome position are illustrated for six ortholog pairs in Figure 7F–G and Figure S6B–C.

To investigate how PATCs may be generated and/or maintained, we analyzed a large set of single nucleotide variants (SNVs) identified in chemically mutagenized and natural variant strains of *C. elegans* (Thompson et al., 2013) (Figure S7). For PATC regions, the underlying periodicity can be used to associate each SNV with a phase (0–10 using A_n/T_n cluster starts as defining phase 0; Fire et al., 2006). Normalized mutation frequencies that result in a net loss of G/C content (G -> A & C -> T transitions and G->T and C->A transversions) within PATCs are most frequent where the G/C content is lowest. In contrast, mutations that increase G/C content are enriched in areas with already high G/C content. These two mutational profiles are offset by approximately 5 basepairs, yielding a tendency to strengthen the density of PATCs within regions already contained in A_n/T_n-rich clusters (Figure S7). These results are consistent with models (e.g. Holmquist, 1994) in which large-scale genomic domains in nematodes are relatively fixed and genes adapt under evolutionary pressure or due to net mutagenic bias to the chromatin domains they are embedded in.

In sum, changes in non-coding DNA reveal a surprisingly dynamic character (on an evolutionary time-scale) with PATCs generated or eliminated and concomitant intron expansion or contraction for genes expressed in the germline.

Discussion

In many cases, epigenetic control of gene expression is initiated by small RNAs and maintained by proteins bound to DNA. Here we show that distributed sequences of non-coding DNA can safeguard genes from epigenetic gene silencing. Specifically, these DNA structures are comprised of clusters of A/T sequences arranged so that they are on a single face of the DNA molecule. These periodic A_n/T_n clusters (PATCs) are found within introns or in intergenic regions and can promote germline expression of transgenes in repressive environments. We propose that PATCs similarly promote expression of endogenous genes from repressive chromatin environments.

Genome domains and position effect variegation in *C. elegans*

Using a synthetic transposon, we probed the *C. elegans* genome for chromatin environments affecting transgene expression. The chromatin domains for transgene expression described here are consistent with the large-scale structural stratification observed for other genome features:

i. *C. elegans* autosomes are partitioned into broad central and distal regions. Here we present functional evidence that these broad genomic domains influence gene expression. Transgenes inserted into distal autosomal regions are frequently silenced in the germline and subject to position effect variegation in somatic cells. Consistent with alternating regions of repressive and permissive chromatin marks (e.g. Gu and Fire, 2010; Liu et al., 2010) transgene silencing on arms is not uniform but rather a heterogeneous mix of active and silenced insertions.

ii. The *C. elegans* X-chromosome is largely inactivated in early meiotic stages of the gonad (Fong et al., 2002; Kelly et al., 2002). We observed expression that closely mimicked this pattern: most X chromosome

insertions were silenced in the early germline. There were two interesting exceptions. First, two P*pie-1* insertions near the left tip of the chromosome were expressed. The left tip of the X chromosome exhibits several features characteristic of distal regions of autosomes and is distinct from the remainder of the X chromosome (Fire et al., 2006; Fong et al., 2002; Kelly et al., 2002). Second, P*smu-1* insertions anywhere on the X chromosome were initially silenced but de-silenced over a few generations. Genes on the X-chromosome are largely devoid of PATCs, which may reflect the unique constraints on X-linked genes caused by X chromosome inactivation in the germline and somatic sex dosage compensation (Meyer, 2010). These data suggest that the *smu-1*:GFP transgene, possibly due to PATCs, can overcome chromosome-scale Polycomb-mediated epigenetic silencing.

A role for an abundant class of non-coding DNA in nematodes

An analysis of the large RNA polymerase II gene (*ama-1*) identified periodic sequence features (A-tracts), which were predicted to result in significant DNA bending and explain the unusual migration of *C. elegans* genomic DNA on electrophoretic gels (VanWye et al., 1991). Subsequent analysis on the completed *C. elegans* genome sequence identified a pervasive periodic 10 bp motif of A_n/T_n clusters that were associated with germline expressed genes (Fire et al., 2006). Here we present experimental support for how genes can acquire PATCs in repressive chromatin environments through biased mutations and evidence supporting a causal role for PATCs in permitting germline expression. Transgenes containing promoters with PATCs (P*pie-1*:GFP and P*mex-5*:GFP) were less prone to silencing than a transgene with few PATCs in the promoter (P*dpy-30*:GFP). PATC distributed throughout the entire construct (*smu-1*:GFP) or only in the coding region (various GFPs) significantly reduced germline silencing compared to transgenes lacking PATCs. These effects were consistent across different PATCs: *C. elegans* and *C. briggsae* introns, synthetic introns, short and long introns, and from a multitude of different genes. Our data support a generally permissive role for PATCs in allowing germline expression in contrast to an instructive signal that directly drives germline expression; we imagine that distributed PATCs within a gene allow “proper” DNA access for transcriptional regulation (enhancers, promoters, transcriptional elongation, splicing, etc.). Importantly, transgenes with many internal PATCs are not immune from stochastic silencing; once silenced, we could not distinguish a PATC-rich transgene from a PATC-poor transgene based on fluorescence imaging, smFISH and RNA sequencing.

PATCs may protect endogenous genes from silencing

Why are PATCs necessary? We have studied the role of PATCs in the context of transgenes but we propose that their natural role is to protect endogenous genes from otherwise non-discriminating silencing mechanisms. There are several described mechanisms by which transgenes are silenced in *C. elegans*: histone methylation, RNAi, RNA epigenetic (RNAe), and unpaired chromosomes in meiosis (Kelly et al., 2002; Ketting et al., 1999; Leopold et al., 2015; Shirayama et al., 2012; Tabara et al., 1999). PATCs may function in parallel to protective pathways that are proposed to depend on a balance between silencing piRNAs

bound to the PRG-1 Argonaute and activating small RNAs bound to the Argonaute protein CSR-1 (Shirayama et al., 2012). A balance between permissive and repressive pathways does not imply full activation or complete silencing; for example, expression of a P*mex-5*::GFP transgene was increased five-fold in *prg-1* mutants (Leopold et al., 2015). More generally, a class of endogenous genes with complementarity to 22G small RNAs were derepressed in *prg-1* mutants, suggesting that piRNAs do not exclusively repress foreign DNA (Lee et al., 2012). Unless mechanisms to "reboot" expression exist, antagonistic repressive and permissive pathways that monitor expression in prior generations would be prone to enter a negative feedback loop, which would ultimately lead to a silenced state. Endogenous structures that counteract silencing, such as PATCs, may constitute a fail-safe mechanism to prevent negative feedback loops from forming. PATCs may be particularly important for endogenous genes located in repressive genomic environments, such as the autosomal arms, where unimpeded spreading of repressive histone marks from adjacent regions may bias the balance toward the repressive pathway. Since no homologs of canonical boundary elements or insulators have been identified in *C. elegans* (Heger et al., 2009), one possible role for distributed PATCs is to prevent heterochromatin spreading.

PATCs may defend the genome against viral or transposon DNA

What are the possible benefits to nematodes for evolving PATC structures? We propose that this unusual DNA structure may be an important component of a genomic immune system that protects the nematode from viral or transposon invasion. A cellular immune system must recognize self versus non-self. We suggest a model in which PATCs can protect endogenous genes from silencing by counteracting silencing pathways. For endogenous genes, the PATC signature can be incorporated into non-coding DNA with no effect on protein sequence. For such a defense system to be efficient, insertion into a PATC-rich chromatin domain should not confer anti-silencing properties; this requirement is consistent with the observed broad distribution of PATCs in endogenous genes and silencing of transgenes inserted into local PATC-rich domains. In the constant arms race between hosts and parasitic DNA elements, it would be costly for invasive DNA to evolve structures that mimic PATCs to avoid silencing, especially if PATCs are specific to nematodes. This protection of self may have allowed nematodes to evolve an aggressive defense system that includes large, repressive genome domains and a piRNA system that tolerates target mismatches. In this model, invasive foreign DNA lacking an elaborate structural feature of endogenous genes are silenced by default (even in central, less repressive regions of the genome) and uncontrolled propagation of invasive DNA is limited. Notably, such a genomic defense system would not require a cellular memory of prior exposure to any particular foreign DNA sequence or structure.

Experimental Procedures

Transgene insertions

Random miniMos insertions were generated in *unc-119(ed3)* animals and targeted mosSCI insertions in animals with Mos1 elements at defined locations (Frøkjær-Jensen et al., 2008, 2014). See Table S1 for strains and insertion sites.

Imaging

Germline fluorescence was classified by visual inspection on a fluorescence dissection microscope blind to strain identity (Table S2). Automated imaging of somatic fluorescence was performed on confocal microscopy images of fixed L1 animals (Liu et al., 2009) or mixed stage profiling of live animals on a COPAS-profiler2 (Dupuy et al., 2007) (Table S2). smFISH was performed according to the manufacturer's protocol (Biosearch Technologies, CA) on dissected germlines, imaged at 100x magnification and deconvolved. mRNA and transcriptional foci were quantified blind to strain identity.

Molecular Biology

All plasmids were generated by standard techniques and annotated Genbank sequences are included in Data S1.

Total mRNA and small RNA sequencing

We isolated mRNA and small RNAs from animals grown at 25°C and sequenced libraries on a miSeq instrument (Illumina, CA). We aligned reads to *C. elegans* protein coding sequences (WS245). SRA accession: SRP072711.

PATC analysis

Individual genes and introns (Table S3) were analyzed with the original PATC algorithm (Fire et al., 2006) and whole genomes were analyzed with a modified algorithm that minimizes off-helical A_n/T_n signals (Table S4). See Data S2 for 25bp resolution PATC signals for *C. elegans* (WS245) and *C. briggsae* (WS245) genomes and Data S3 for intron sequences incorporated in *gfp*.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank J. Feldman, A. Villeneuve, S. Strome, A. Rechtsteiner, J. Lieb, K. Ikegami, C. Engert, S. Klemm, T. Machacek, V. Jantsch, C. Girard, G. Maro, D. Nix, Fire and Jorgensen lab members; and K. Hoe for expert technical assistance. This work was supported by the Carlsberg Foundation (C.F.J), Direktør Ib Henriksens Foundation (C.F.J.), a Stanford Graduate Fellowship (N.J.), National Institutes of Health grants R01GM095817 (E.M.J) and R01GM37706 (A.Z.F.), and the Howard Hughes Medical Institute (E.M.J.).

References

Aravin AA, Hannon GJ, Brennecke J. The Piwi-piRNA Pathway Provides an Adaptive Defense in the Transposon Arms Race. Science. 2007; 318:761–764. [PubMed: 17975059]

Batista PJ, Ruby JG, Claycomb JM, Chiang R, Fahlgren N, Kasschau KD, Chaves DA, Gu W, Vasale JJ, Duan S, et al. PRG-1 and 21U-RNAs Interact to Form the piRNA Complex Required for Fertility in C. elegans. Mol Cell. 2008; 31:67–78. [PubMed: 18571452]

Brenner S. The Genetics of Caenorhabditis Elegans. Genetics. 1974; 77:71–94. [PubMed: 4366476]

Das PP, Bagijn MP, Goldstein LD, Woolford JR, Lehrbach NJ, Sapetschnig A, Buhecha HR, Gilchrist MJ, Howe KL, Stark R, et al. Piwi and piRNAs Act Upstream of an Endogenous siRNA Pathway to Suppress Tc3 Transposon Mobility in the Caenorhabditis elegans Germline. Mol Cell. 2008; 31:79–90. [PubMed: 18571451]

Dumesic PA, Natarajan P, Chen C, Drinnenberg IA, Schiller BJ, Thompson J, Moresco JJ, Yates JR, Bartel DP, Madhani HD. Stalled spliceosomes are a signal for RNAi-mediated genome defense. Cell. 2013; 152:957–968. [PubMed: 23415457]

Dupuy D, Bertin N, Hidalgo CA, Venkatesan K, Tu D, Lee D, Rosenberg J, Svrzikapa N, Blanc A, Carnec A, et al. Genome-scale analysis of in vivo spatiotemporal promoter activity in Caenorhabditis elegans. Nat Biotechnol. 2007; 25:663–668. [PubMed: 17486083]

Elgin SCR, Reuter G. Position-effect variegation, heterochromatin formation, and gene silencing in Drosophila. Cold Spring Harb Perspect Biol. 2013; 5:a017780. [PubMed: 23906716]

Fire A, Alcazar R, Tan F. Unusual DNA structures associated with germline genetic activity in Caenorhabditis elegans. Genetics. 2006; 173:1259–1273. [PubMed: 16648589]

Fong Y, Bender L, Wang W, Strome S. Regulation of the different chromatin states of autosomes and X chromosomes in the germ line of C. elegans. Science. 2002; 296:2235–2238. [PubMed: 12077420]

Frøkjær-Jensen C, Wayne Davis M, Hopkins CE, Newman BJ, Thummel JM, Olesen S-P, Grunnet M, Jorgensen EM. Single-copy insertion of transgenes in Caenorhabditis elegans. Nat Genet. 2008; 40:1375–1383. [PubMed: 18953339]

Frøkjær-Jensen C, Davis MW, Sarov M, Taylor J, Flibotte S, LaBella M, Pozniakovsky A, Moerman DG, Jorgensen EM. Random and targeted transgene insertion in Caenorhabditis elegans using a modified Mos1 transposon. Nat Methods. 2014; 11:529–534. [PubMed: 24820376]

Garrigues JM, Sidoli S, Garcia BA, Strome S. Defining heterochromatin in C. elegans through genome-wide analysis of the heterochromatin protein 1 homolog HPL-2. Genome Res. 2015; 25:76–88. [PubMed: 25467431]

Gu SG, Fire A. Partitioning the C. elegans genome by nucleosome modification, occupancy, and positioning. Chromosoma. 2010; 119:73–87. [PubMed: 19705140]

Heger P, Marin B, Schierenberg E. Loss of the insulator protein CTCF during nematode evolution. BMC Mol Biol. 2009; 10:84. [PubMed: 19712444]

Hirsh D, Oppenheim D, Klass M. Development of the reproductive system of Caenorhabditis elegans. Dev Biol. 1976; 49:200–219. [PubMed: 943344]

Ho JWK, Jung YL, Liu T, Alver BH, Lee S, Ikegami K, Sohn KA, Minoda A, Tolstorukov MY, Appert A, et al. Comparative analysis of metazoan chromatin organization. Nature. 2014; 512:449–452. [PubMed: 25164756]

Ikegami K, Egelhofer TA, Strome S, Lieb JD. Caenorhabditis elegans chromosome arms are anchored to the nuclear membrane via discontinuous association with LEM-2. Genome Biol. 2010; 11:R120. [PubMed: 21176223]

Johnson CL, Spence AM. Epigenetic licensing of germline gene expression by maternal RNA in C. elegans. Science. 2011; 333:1311–1314. [PubMed: 21885785]

Kelly WG, Xu S, Montgomery MK, Fire A. Distinct requirements for somatic and germline expression of a generally expressed Caernorhabditis elegans gene. Genetics. 1997; 146:227–238. [PubMed: 9136012]

Kelly WG, Schaner CE, Dernburg AF, Lee M-H, Kim SK, Villeneuve AM, Reinke V. X-chromosome silencing in the germline of C. elegans. Dev Camb Engl. 2002; 129:479–492.

Ketting RF, Haverkamp TH, van Luenen HG, Plasterk RH. Mut-7 of C. elegans, required for transposon silencing and RNA interference, is a homolog of Werner syndrome helicase and RNaseD. Cell. 1999; 99:133–141. [PubMed: 10535732]

Lee HC, Gu W, Shirayama M, Youngman E, Conte D Jr, Mello CC. C. elegans piRNAs mediate the genome-wide surveillance of germline transcripts. Cell. 2012; 150:78–87. [PubMed: 22738724]

Leopold LE, Heestand BN, Seong S, Shtessel L, Ahmed S. Lack of pairing during meiosis triggers multigenerational transgene silencing in Caenorhabditis elegans. Proc Natl Acad Sci. 2015; 112:E2667–E2676. [PubMed: 25941370]

Liu T, Rechtsteiner A, Egelhofer TA, Vielle A, Latorre I, Cheung M-S, Ercan S, Ikegami K, Jensen M, Kolasinska-Zwierz P, et al. Broad chromosomal domains of histone modification patterns in C. elegans. Genome Res. 2010

Liu X, Long F, Peng H, Aerni SJ, Jiang M, Sánchez-Blanco A, Murray JI, Preston E, Mericle B, Batzoglou S, et al. Analysis of cell fate from single-cell gene expression profiles in C. elegans. Cell. 2009; 139:623–633. [PubMed: 19879847]

MacQueen AJ, Phillips CM, Bhalla N, Weiser P, Villeneuve AM, Dernburg AF. Chromosome Sites Play Dual Roles to Establish Homologous Synapsis during Meiosis in C. elegans. Cell. 2005; 123:1037–1050. [PubMed: 16360034]

Malone CD, Hannon GJ. Small RNAs as guardians of the genome. Cell. 2009; 136:656–668. [PubMed: 19239887]

Meister P, Towbin BD, Pike BL, Ponti A, Gasser SM. The Spatial Dynamics of Tissue-Specific Promoters During C. Elegans Development. Genes Dev. 2010; 24:766–782. [PubMed: 20395364]

Meyer BJ. Targeting X chromosomes for repression. Curr Opin Genet Dev. 2010; 20:179–189. [PubMed: 20381335]

Ortiz MA, Noble D, Sorokin EP, Kimble J. A New Dataset of Spermatogenic vs. Oogenic Transcriptomes in the Nematode Caenorhabditis elegans. G3 GenesGenomesGenetics. 2014; 4:1765–1772.

Raj A, van den Bogaard P, Rifkin SA, van Oudenaarden A, Tyagi S. Imaging individual mRNA molecules using multiple singly labeled probes. Nat Methods. 2008; 5:877–879. [PubMed: 18806792]

Rechtsteiner A, Ercan S, Takasaki T, Phippen TM, Egelhofer TA, Wang W, Kimura H, Lieb JD, Strome S. The Histone H3K36 Methyltransferase MES-4 Acts Epigenetically to Transmit the Memory of Germline Gene Expression to Progeny. PLoS Genet. 2010; 6:e1001091. [PubMed: 20824077]

Rockman MV, Kruglyak L. Recombinational landscape and population genomics of Caenorhabditis elegans. PLoS Genet. 2009; 5:e1000419. [PubMed: 19283065]

Ross JA, Koboldt DC, Staisch JE, Chamberlin HM, Gupta BP, Miller RD, Baird SE, Haag ES. Caenorhabditis briggsae Recombinant Inbred Line Genotypes Reveal Inter-Strain Incompatibility and the Evolution of Recombination. PLoS Genet. 2011; 7:e1002174. [PubMed: 21779179]

Shirayama M, Seth M, Lee HC, Gu W, Ishidate T, Conte D Jr, Mello CC. piRNAs initiate an epigenetic memory of nonself RNA in the C. elegans germline. Cell. 2012; 150:65–77. [PubMed: 22738726]

Spartz AK, Herman RK, Shaw JE. SMU-2 and SMU-1, Caenorhabditis elegans homologs of mammalian spliceosome-associated proteins RED and fSAP57, work together to affect splice site choice. Mol Cell Biol. 2004; 24:6811–6823. [PubMed: 15254247]

Spike CA, Shaw JE, Herman RK. Analysis of smu-1, a gene that regulates the alternative splicing of unc-52 pre-mRNA in Caenorhabditis elegans. Mol Cell Biol. 2001; 21:4985–4995. [PubMed: 11438655]

Stoeckius M, Grün D, Kirchner M, Ayoub S, Torti F, Piano F, Herzog M, Selbach M, Rajewsky N. Global characterization of the oocyte-to-embryo transition in Caenorhabditis elegans uncovers a novel mRNA clearance mechanism. EMBO J. 2014; 33:1751–1766. [PubMed: 24957527]

Strome S, Powers J, Dunn M, Reese K, Malone CJ, White J, Seydoux G, Saxton W. Spindle Dynamics and the Role of γ-Tubulin in Early Caenorhabditis elegans Embryos. Mol Biol Cell. 2001; 12:1751–1764. [PubMed: 11408582]

Tabara H, Sarkissian M, Kelly WG, Fleenor J, Grishok A, Timmons L, Fire A, Mello CC. The rde-1 Gene, RNA Interference, and Transposon Silencing in C. elegans. Cell. 1999; 99:123–132. [PubMed: 10535731]

Thompson O, Edgley M, Strasbourger P, Flibotte S, Ewing B, Adair R, Au V, Chaudhry I, Fernando L, Hutter H, et al. The million mutation project: a new approach to genetics in Caenorhabditis elegans. Genome Res. 2013; 23:1749–1762. [PubMed: 23800452]

VanWye JD, Bronson EC, Anderson JN. Species-specific patterns of DNA bending and sequence. Nucleic Acids Res. 1991; 19:5253–5261. [PubMed: 1923808]

Volpe TA, Kidner C, Hall IM, Teng G, Grewal SIS, Martienssen RA. Regulation of Heterochromatic Silencing and Histone H3 Lysine-9 Methylation by RNAi. Science. 2002; 297:1833–1837. [PubMed: 12193640]

Wang X, Zhao Y, Wong K, Ehlers P, Kohara Y, Jones SJ, Marra MA, Holt RA, Moerman DG, Hansen D. Identification of genes expressed in the hermaphrodite germ line of C. elegans using SAGE. BMC Genomics. 2009; 10:213. [PubMed: 19426519]

Zamore PD, Tuschl T, Sharp PA, Bartel DP. RNAi: Double-Stranded RNA Directs the ATP-Dependent Cleavage of mRNA at 21 to 23 Nucleotide Intervals. Cell. 2000; 101:25–33. [PubMed: 10778853]

Article Highlights

- Periodic, non-coding DNA can prevent transgenes from stochastic silencing in germline
- Non-coding content of genes is shaped by genomic context and heterochromatin domains
- Conditioning of active DNA may allow cells to distinguish foreign from host genes

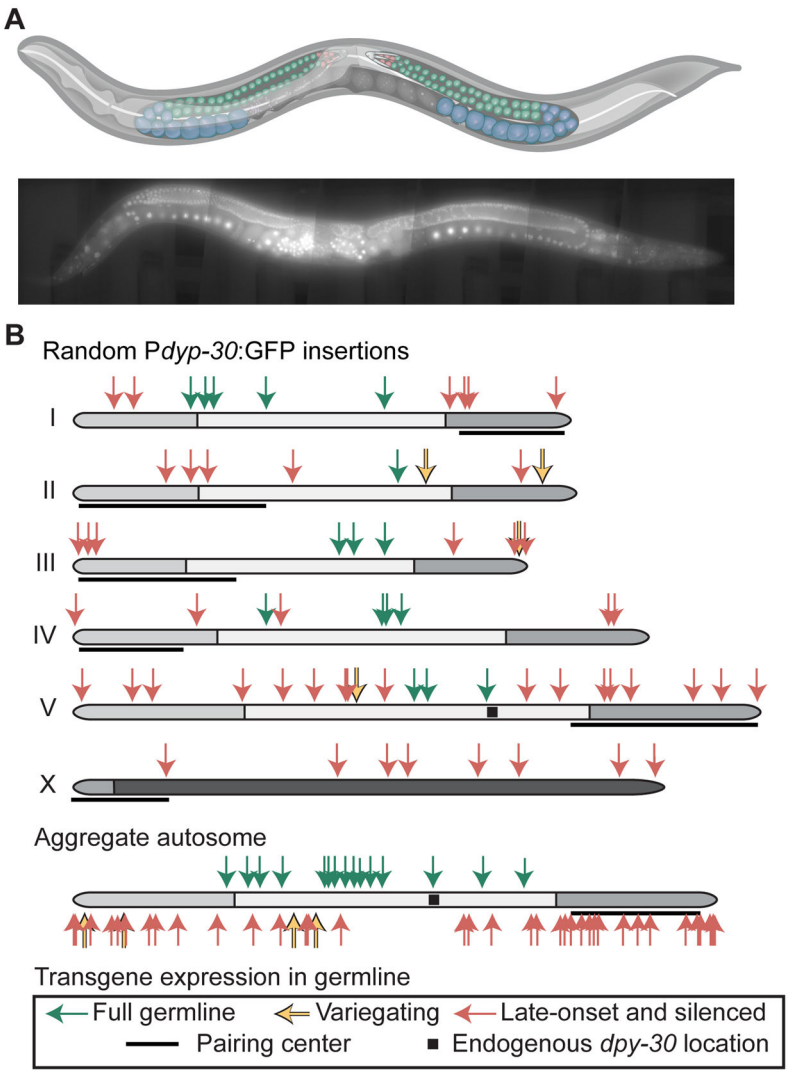


Figure 1. Many P*dpy-30*:GFP transgenes are specifically silenced in the germline

A. Top. Schematic of *C. elegans* hermaphrodite with the female germline highlighted. Red = mitotic cells; green = early meiotic cells; blue = late meiotic cells. Bottom. Composite fluorescence image of animal expressing a P*dpy-30*:GFP transgene (42x magnification, scale bar = 50 micron). Graphic of *C. elegans* modified from "*Caenorhabditis elegans* hermaphrodite adult-en.sv" by K.D. Schroeder from Wikimedia Commons under a CC-BY-SA 3.0 license.

B. Top. Germline expression at 25°C of P*dpy-30*:GFP transgenes inserted randomly by Mos1 transposition. Genomic insertion sites, transgene copy number, and somatic expression of all insertions were previously verified (Frøkjær-Jensen et al., 2014). Germline fluorescence is indicated by colored arrows (key at bottom. Black squares indicate the endogenous location of *dpy-30*, gray shading indicates enrichment of repressive histone modifications (Liu et al., 2010), and pairing centers are indicated with black lines (MacQueen et al., 2005). Bottom. Aggregated normalized autosomes aligned with pairing centers to the right and P*dpy-30*:GFP insertions.

See also Figure S1 and Table S2 for details of classification.

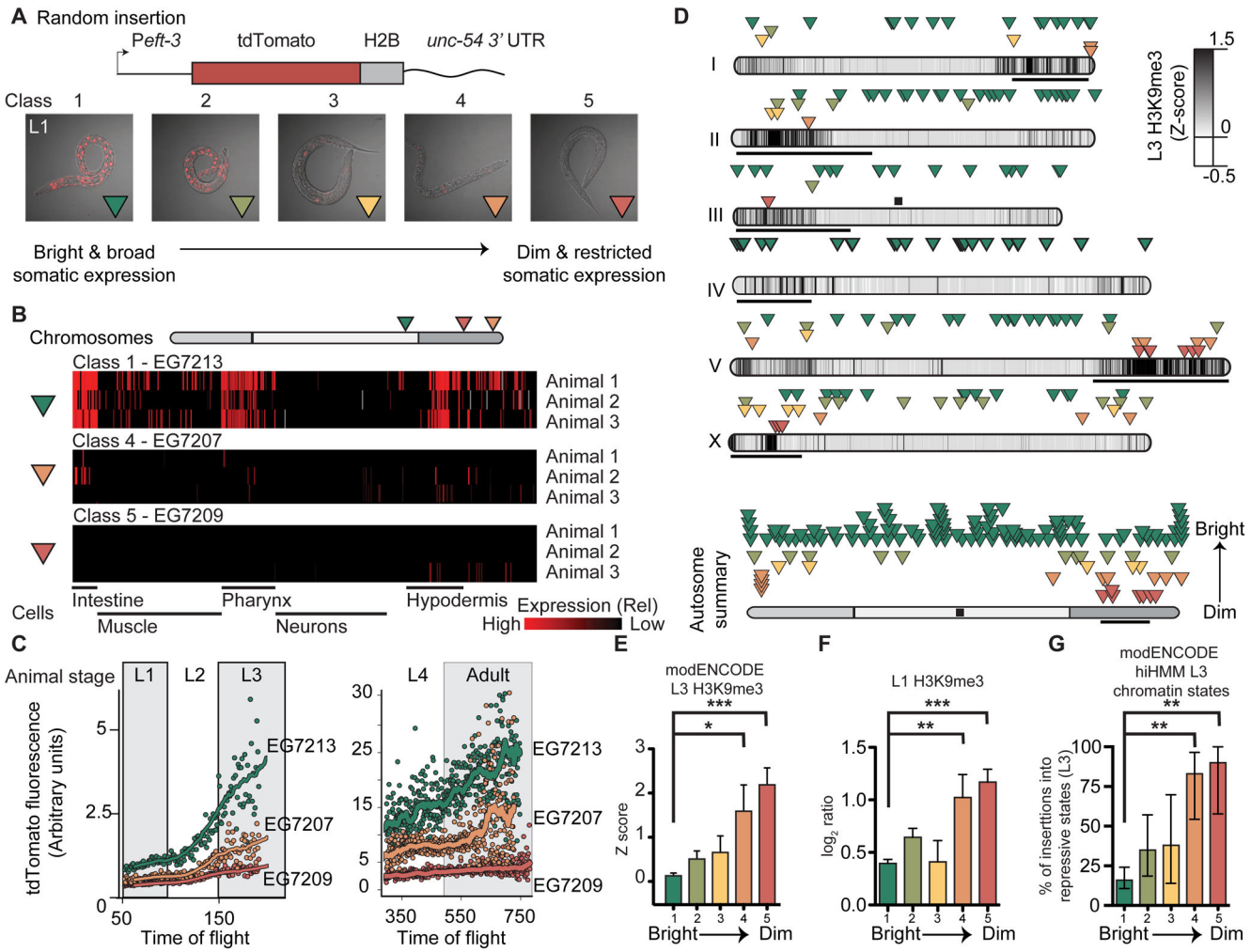


Figure 2. Somatic transgene silencing is frequent on chromosome arms

A. Top. Schematic of the P*eft-3:tdTomato* transgene. Bottom. Overlay of tdTomato fluorescence and transmitted light images (identical exposure) of representative L1 animals from each expression class.

B. Automated analysis of single cell fluorescence in L1 animals (Liu et al., 2009) of one bright and two dimmer P*eft-3:tdTomato:h2b* insertions. Each horizontal line represents replicate imaging of different animals from the same strain. The brightness of each red vertical line indicates the level of tdTomato fluorescence in individual, identified cells. Gray indicates failed cell identification. The cells are clustered based on general classes of tissue and indicated below.

C. tdTomato fluorescence intensity at different larval stages (L1 to adult) based on flow cytometry (Dupuy et al., 2007). Average peak tdTomato fluorescence is indicated for each animal with a dot and the line indicates a smoothed average of peak fluorescence.

D. Top. Genomic location and average brightness of each strain (Green: bright & broad expression, red: dim and restricted expression). Darker chromosome shading indicates higher H3K9me3 density at the L3 stage (Liu et al., 2010), black bars indicate the approximate location of pairing centers (MacQueen et al., 2005), and the black box indicates

the endogenous location of *eft-3*. Bottom. Aggregate normalized autosome with all insertions.

E. Average H3K9me3 Z score in L3 animals (Liu et al., 2010) in a 2 kb interval centered on insertions from each class.

F. Average H3K9me3 level (this work) in starved L1 animals in a 2 kb interval centered on insertions from each class. Log2 ratio between sample and input.

Panels E–F: Average ± SEM. Statistics: Mann-Whitney test (** P< 0.01, *** P < 0.001).

G. Percentage of inserts from each class (Class 1 – Class 5) that was inserted into repressive chromatin states (Polycomb or heterochromatin, states 10–13) identified at the L3 stage based on hierarchical non-parametric machine-learning (hiHMM) (Ho et al., 2014). Error bars indicate 95% confidence interval. Statistics: Fischer's exact test of Class 1 (brightest) compared to Classes 2–5 (dimmer) (** P< 0.01).

See also Figure S2 and Table S2 for higher resolution images and expression quantification.

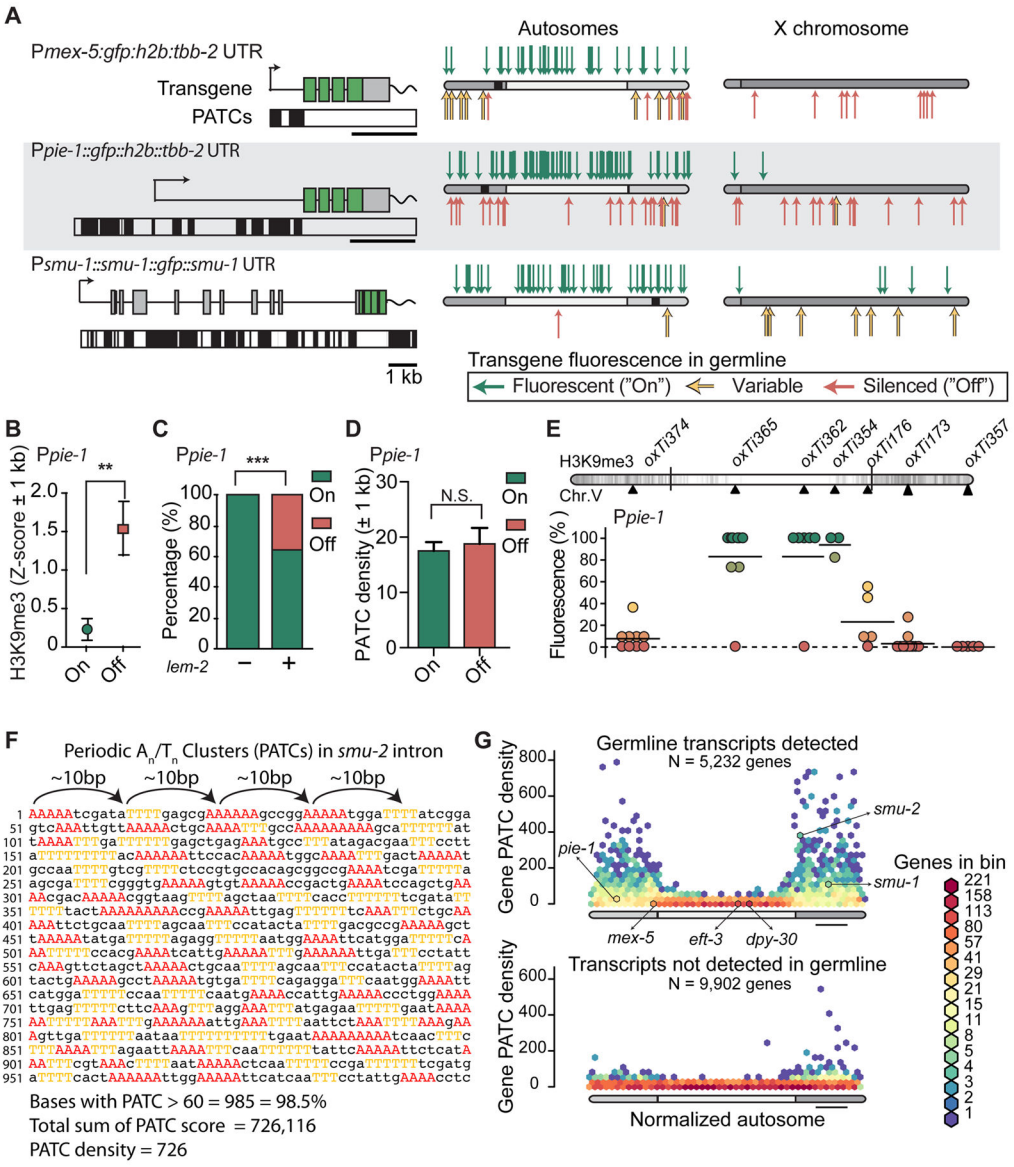


Figure 3. Transgenes containing PATCs are less frequently silenced

A. Expression of P*pie-1:gfp*, P*mex-5:gfp* and *smu-1:gfp* transgenes. Left: transgene schematics with PATCs >60 indicated below as black boxes. Right. Location and germline expression of insertions on aggregated autosomes and the X-chromosome. Endogenous locations of *pie-1*, *mex-5*, and *smu-1* are indicated with black squares.

B. Local chromatin environment (2kb interval centered on insertion sites) near P*pie-1* insertions. H3K9me3 signal from early embryos (Liu et al., 2010), which have been used as a proxy for germline tissue (Rechtsteiner et al., 2010). Mean ± SEM. Statistical test: Mann Whitney (** = P < 0.01).

C. Local chromatin interactions with nuclear lamin (2 kb interval) near P*pie-1* insertions. Nuclear lamin interactions based on ChIP-sequencing on mixed stage embryos with an antibody against the transmembrane nuclear protein *lem-2* (Ikegami et al., 2010). Mean ± SEM. Fischer's exact test (*** = P < 0.001).

D. Local PATC density (2 kb interval) near P*pie-1* insertions. Mean ± SEM. Statistical test: Mann Whitney (N.S. = not significant).
E. Germline expression of targeted, single-copy P*pie-1*:GFP:H2B:*tbb-2* UTR insertions into universal MosSCI sites on Chr. V (see also Figure S1). Circles indicate independent transgene insertions that are color coded for each insertion's expression in the germline (11 animals scored, horizontal bar = mean of independent insertions). Darker chromosome shades correspond to higher H3K9me3 density in early embryos (Liu et al., 2010).
F. Example of a Periodic A_n/T_n Cluster (PATC) from intron 3 of *smu-2*. Clusters of three, four, or five adjacent As and Ts are colored. Clusters of As and Ts are separated by approximately 10 bps (~one helical DNA turn) and short A_n and T_n clusters therefore align along one face of the DNA helix over an extended region (here 1 kb). The PATC algorithm (Fire et al., 2006) assigns a PATC value to every nucleotide of a DNA sequence; higher values indicate that the nucleotide is part of an extended DNA stretch ("PATC-rich region") with many clusters of A_n/T_n clusters in perfect 10-basepair register. Less than 0.1% of nucleotides in a random DNA sequence reach a PATC value of 60. The PATC density is defined as the average PATC value of nucleotides in a sequence (See (Fire et al., 2006) and Supplemental Information for details)).
G. Comparison of the PATC density of autosomal, protein-coding genes as a function of chromosome position. Top. Germline-expressed genes (based on > 2 FPKM expression in 1-cell oocytes, (Stoeckius et al., 2014)). Bottom. Genes with no detectable germline expression. Gene distributions were hexagonally binned on a logarithmic frequency scale. A subset of genes used in this study are indicated with arrows.
See also Figure S3 and Table S3.

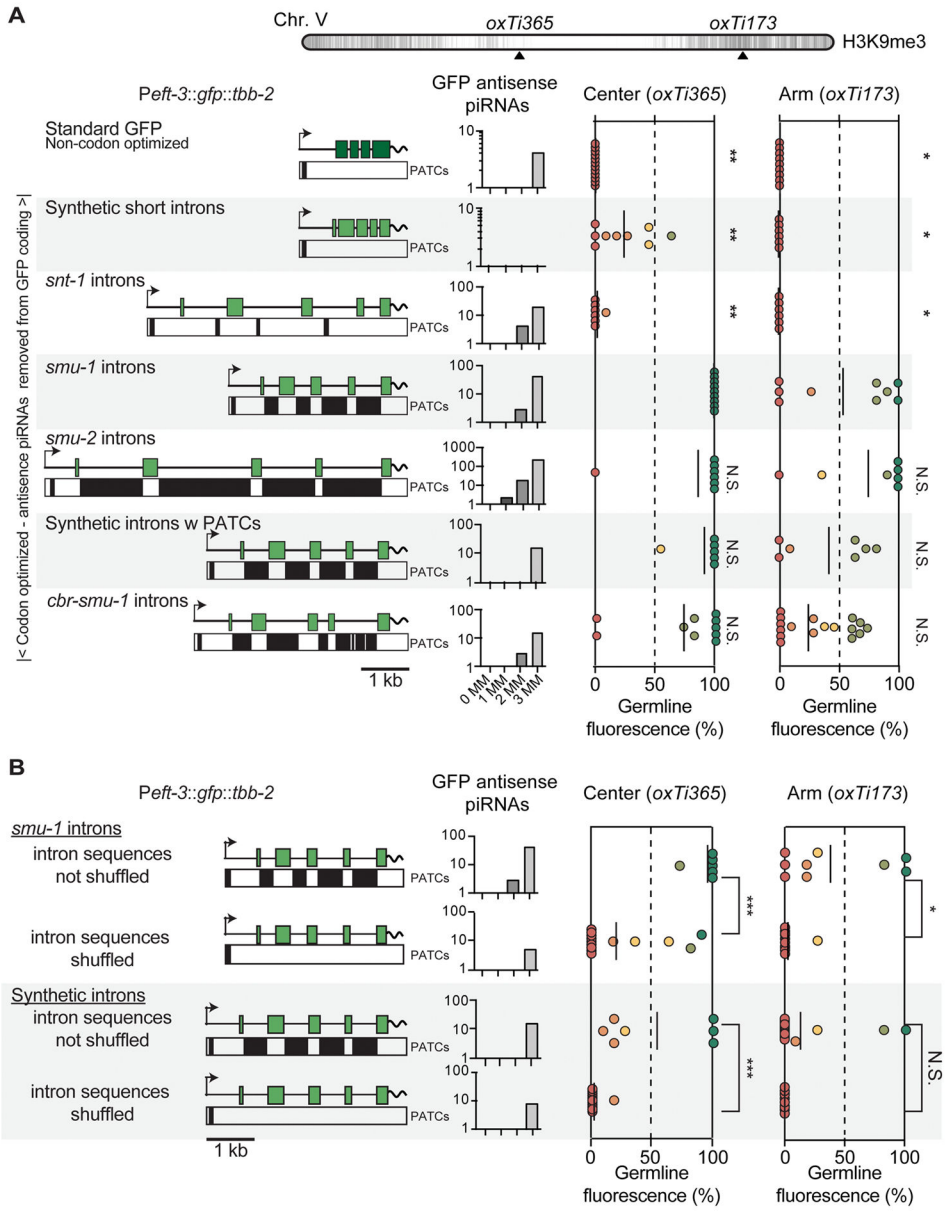


Figure 4. PATCs in GFP introns reduce germline silencing

A. Germline expression of transgenes inserted into chromosome V at 25°C. Top. Location of MosSCI insertion sites on Chr. V. Left, P*eft-3:gfp* transgenes with PATCs >55 indicated as black boxes underneath. Center. piRNA homology of transgenes allowing 0, 1, 2, or 3 mismatches (MM). Right, germline expression from insertions into *oxTi365* and *oxTi173*, respectively. The standard *gfp* (dark green) contains three synthetic introns and was not codon-optimized for *C. elegans* (Fire et al., 1998b). All other *gfp*s (light green) were *C. elegans* codon-optimized and homologies with less than four mismatches to piRNAs were removed. All codon-optimized *gfp* transgenes were identical except for the four introns. “Synthetic introns” and “synthetic introns w PATCs” were generated by gene synthesis. Statistical test: Kruskal-Wallis ANOVA comparing all insertions at a given genomic

location. Dunn's multiple comparison t-test against *gfp* with *smu-1* introns (* P < 0.05, ** P< 0.01).

B. Germline expression of transgenes with semi-randomly shuffled intron sequences (shuffling done with rules to prevent novel consensus splice site motifs while maintaining basepair composition of introns). Statistical test: Mann-Whitney rank test (* P < 0.05, ** P< 0.01, *** P< 0.005, N.S. = not significant).

Panels A–B. All transgenes were expressed in somatic cells. P*eft-3:gfp* transgenes with *smu-1* introns and synthetic PATCs in panel A and panel B are the same transgene constructs; however, independent insertions were generated in parallel to the shuffled transgene insertions for comparison under identical conditions.

See also Figure S4.

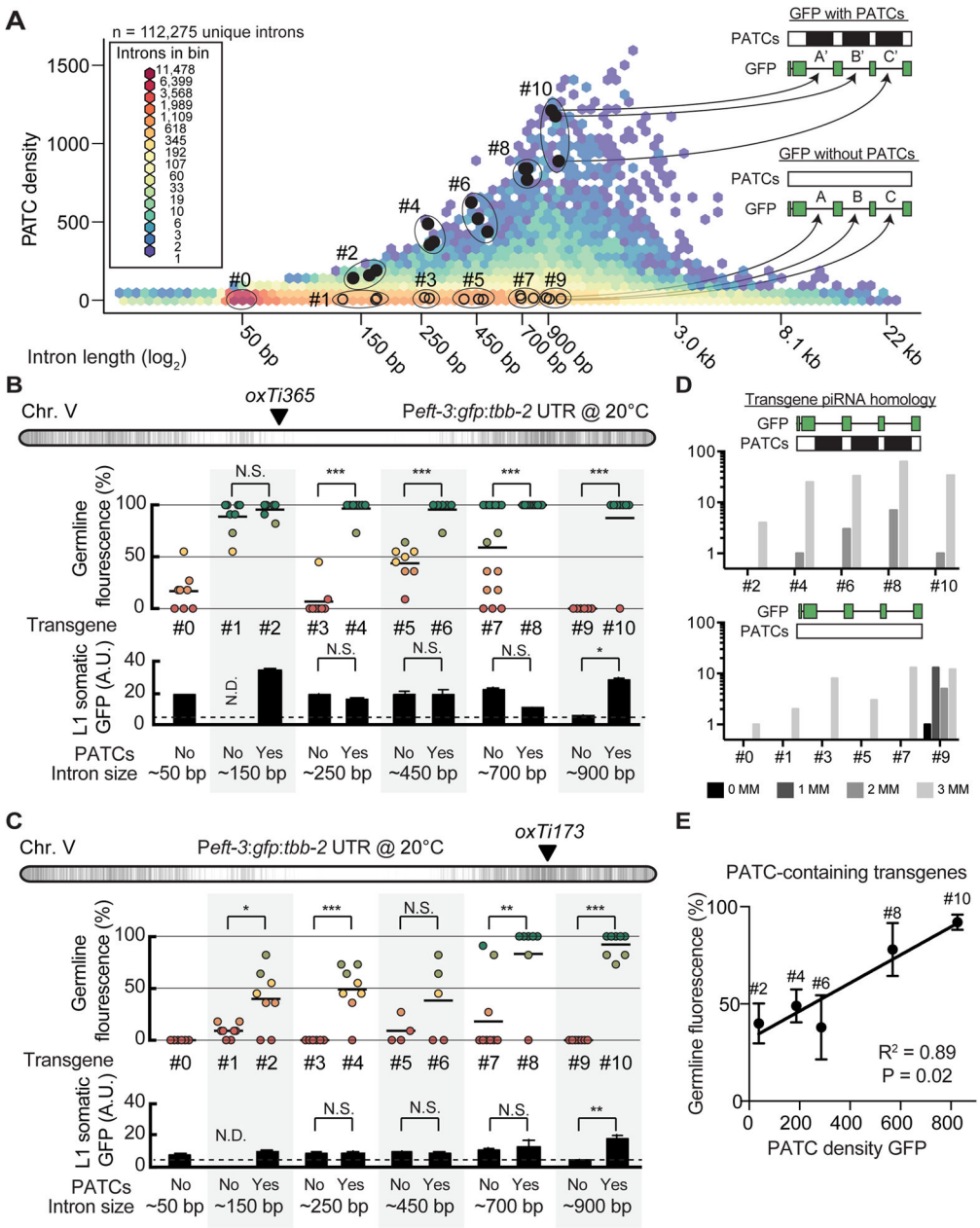


Figure 5. PATC algorithm can identify introns that reduce germline silencing

A. Hexagon plot showing the PATC density of all predicted, unique introns extracted from protein-coding genes (WS245). Introns tested in panel B are indicated by closed black circles (high PATC density) and open circles (low PATC density). Each individual intron with high PATC content (top) was matched to an intron with low PATC content (bottom) based on two parameters: (1) intron length and (2) germline expression of the parent gene (both within 10%).

B. Germline (top) and somatic (bottom) expression at 20°C from single-copy P*eft-3:gfp* transgene insertions at a central, permissive chromosome location (*oxTi365*).

C. Germline (top) and somatic (bottom) expression at 20°C from single-copy P*eft-3:gfp* transgene insertions at a distal, repressive chromosome location (*oxTi173*).
Panel B–C. Statistical test: Mann-Whitney (* P < 0.05, ** P< 0.01, *** P< 0.005, N.S. = not significant, N.D = no data).
D. piRNA homology with zero (0MM), one (1MM), two (2MM), or three (3MM) mismatches to the sequence of each numbered *gfp*.
E. Linear correlation between germline expression and PATC content for PATC-rich transgenes inserted at the repressive *oxTi173* location. The P value indicates the statistical significance for a positive slope of the linear fit.
See also Table S3.

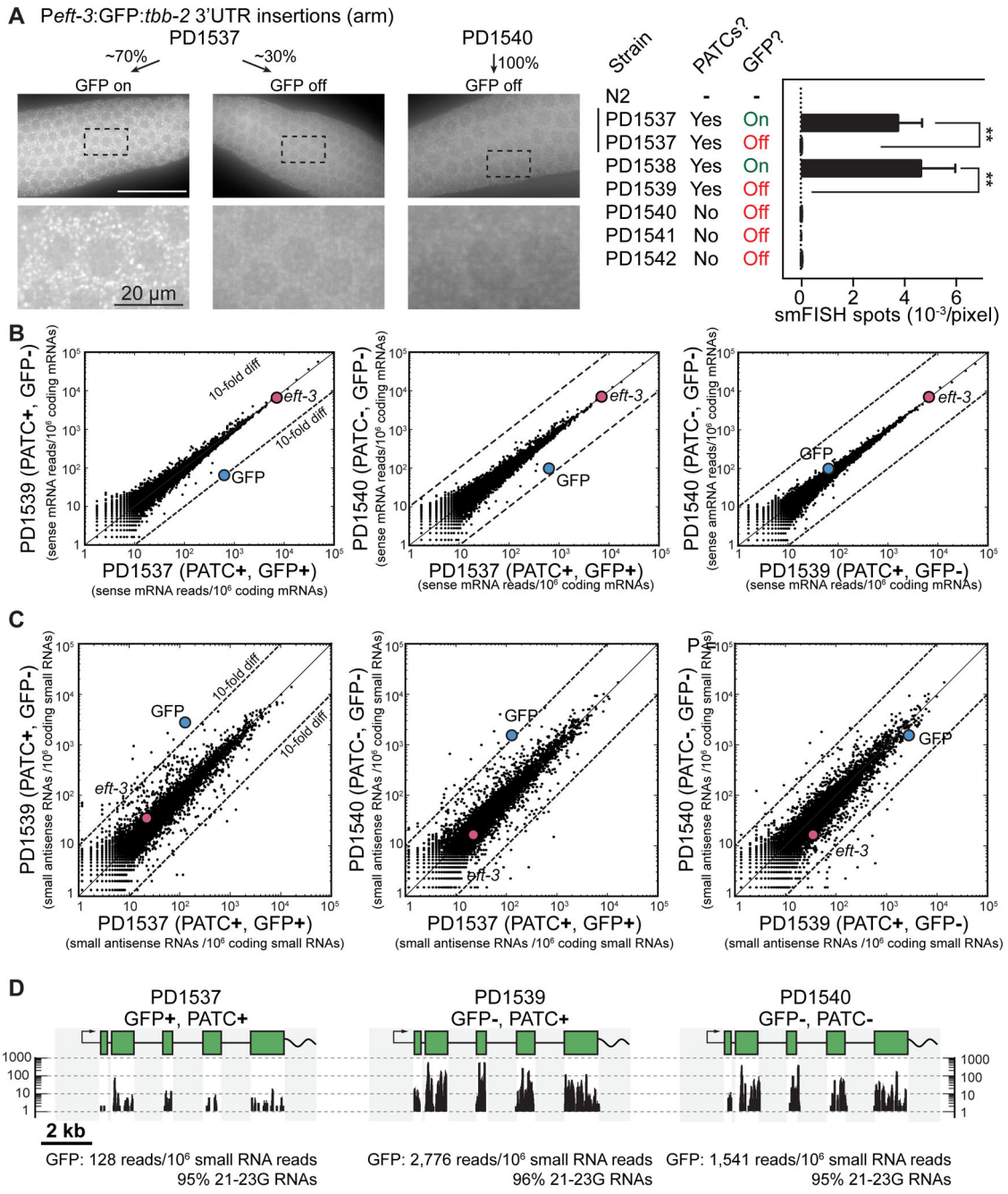


Figure 6. Transgene mRNA expression and small RNA populations

A. Left. Representative images from single molecule Fluorescence *In Situ* Hybridization (smFISH) on the germline from animals with frequent GFP expression (~70% fluorescent germlines, PD1537) and a fully silenced strain (PD1540) from P*eft-3:gfp* transgenes. Individual white spots indicate diffraction-limited single mRNA transcripts (Raj et al., 2008) and germline nuclei are visible as darker circles. Right. Quantification (blinded) of the number of smFISH spots (mean ± SEM) from transgenic animals and N2 animals. In many cases, germlines with silenced *gfp* showed easily distinguishable GFP expression in somatic

cells and smFISH spots in those tissues. Statistical test: ANOVA, post-test Sidak's multiple comparison test (** P< 0.01).

B. Sense mRNA expression of young adult animals with frequent germline expression (PD1537) or complete germline silencing (PD1539, PD1540) of P*eft-3:gfp* transgenes. Statistical tests: 2-proportion Z-test (two-tailed), number of *gfp* transcripts in PD1537 vs PD1539, P ≈ 0; PD1537 vs PD1540, P ≈ 0.

C. Top. Small antisense RNA expression in young adult animals with germline expressed (PD1537) or silenced (PD1539, PD1540) P*eft-3:gfp* transgenes. Statistical tests: 2-proportion Z-test (two-tailed), number of *gfp* transcripts in PD1537 vs PD1539, P ≈ 0; PD1537 vs PD1540, P ≈ 0.

Panel B + C. RNAs were aligned against all protein coding genes (WS245) and normalized to uniquely aligned RNAs. Colored dots indicate *gfp* and *eft-3* expression. All strains had GFP expression in somatic cells.

D. Unique alignments of small RNAs (>95% 21–23G RNAs) detected against *gfp*. The native introns in *gfp* (250 bp), the *eft-3* promoter, and *tbb-2* 3′UTR are also present in the *C. elegans* genome and therefore reads could not be aligned uniquely to these regions (shaded gray).

See also Figure S5.

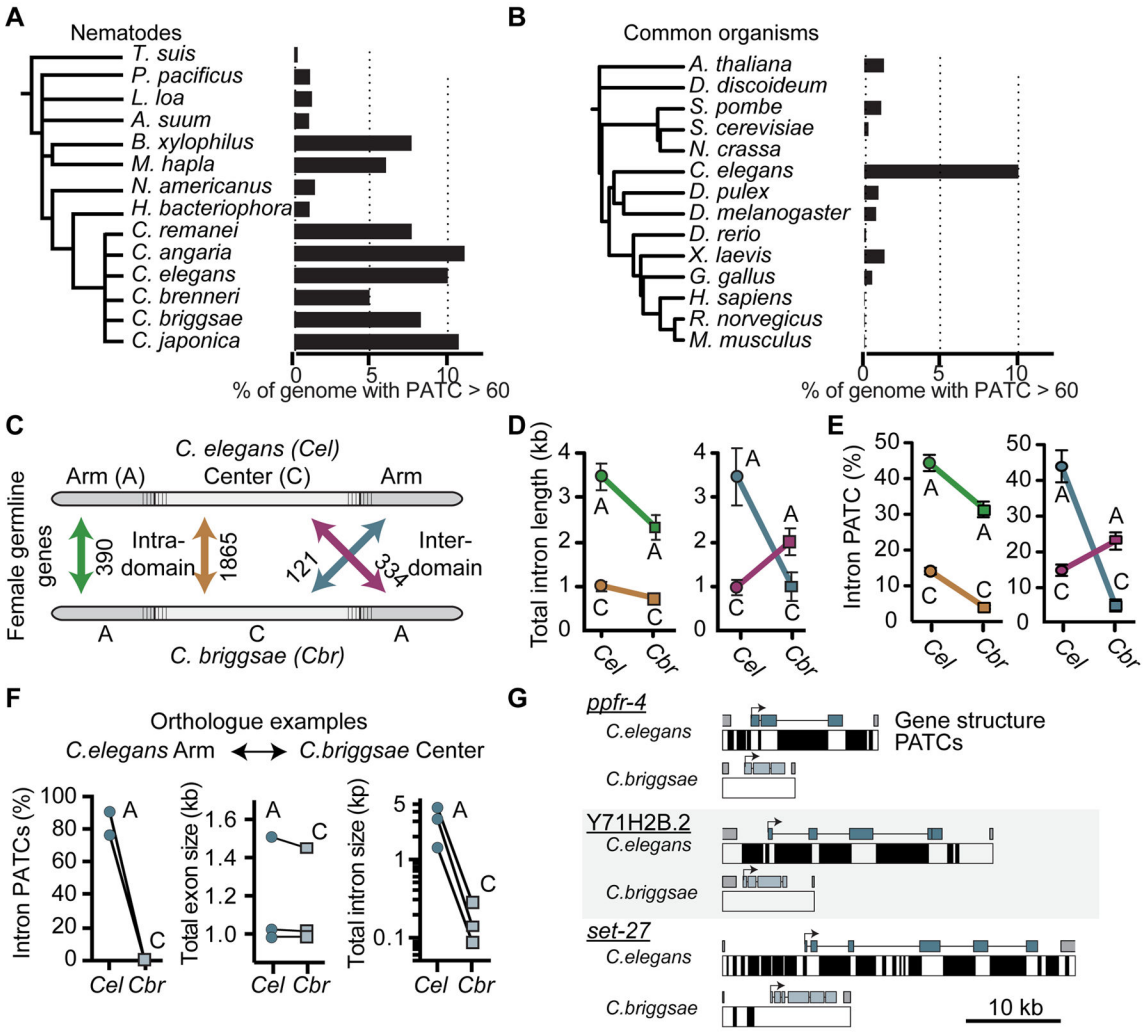


Figure 7. *C. elegans* and *C. briggsae* intron size and PATC content contract and expand with chromosomal location

A. PATC content of select nematode genomes.

B. PATC signal in commonly studied genetic model organisms and the human genome sequence. Panel A and B: frequency of DNA with a PATC value greater than 60.

C. Schematic of the unique orthologs analyzed. One class ("intra-domain") contains orthologs that reside on the arm (green) or the center (brown) of both *C. elegans* and *C. briggsae*. A second class contains "inter-domain" orthologs (blue, red) that have moved between chromatin domains in the two species.

D. Comparison of the total intron length of genes for each ortholog class.

E. Comparison of the percentage of intronic bases with an average PATC > 60 for each ortholog class.

F. An example of the PATC content, exon length, and intron size for three genes residing in a distal repressive chromatin domain in *C. elegans* and a central permissive domain in *C. briggsae* on Chr. III.

G. Gene structure and PATCs > 60 of the three genes from Panel F. The DNA sequence from the last exon of the upstream gene (“5′ gene”) to the first intron of the downstream gene (“3′ gene”) are included.

See also Figures S6–7, Table S4, and Data S2.